The Annals of Statistics
1988, Vol. 16, No. 1, 265–277

SIMULTANEOUS ESTIMATION AND PREDICTION USING THE EXPECTED COVERAGE MEASURE CRITERION[1]

BY PETER M. HOOPER

University of Alberta

Simultaneous confidence regions and simultaneous prediction regions are considered using the expected coverage measure criterion of Naiman. Invariance methods are used to construct minimax procedures in multivariate linear models.

1. Introduction. Let $\mathbf{x}$ and $\mathbf{y}$ be independent random vectors with

$$\mathbf{x} \sim N_n(C_1\beta, \sigma^2 I_n) \quad \text{and} \quad \mathbf{y} \sim N_m(C_2\beta, c_0\sigma^2 I_m), \tag{1.1}$$

where $\beta \in \mathbb{R}^q$, $\sigma > 0$ are unknown and C_1, C_2 and c_0 are known. We observe $\mathbf{x}$, but not $\mathbf{y}$, and we want intervals for $a'\mathbf{y}$, for some vectors $a \in \mathbb{R}^m$. We wish to construct intervals $R(a, \mathbf{x})$ so that, in most cases, $a'\mathbf{y}$ is covered. The intervals should also be narrow.

Naiman (1984) suggests the following approach to balancing these aims. Define a probability measure μ on $\mathbb{R}^m$, a positive weight function $w(a, \sigma)$ and a bound $1 - \alpha \in (0, 1)$. Then choose R to minimize

$$E\int w(a, \sigma) l(R(a, \mathbf{x}))\mu(da) \tag{1.2}$$

subject to

$$E\int I_{R(a,\mathbf{x})}(a'\mathbf{y})\mu(da) \geq 1 - \alpha \quad \text{for all } \beta \in \mathbb{R}^q, \sigma > 0. \tag{1.3}$$

In the preceding, l is Lebesgue measure on $\mathbb{R}$, $I_R(z) = 1$ if $z \in R$ and $I_R(z) = 0$ if $z \notin R$. The minimization of (1.2) could be in the minimax sense or uniform in (β, σ) subject to further restrictions on R. Naiman (1984) calls the left-hand side of (1.3) the *expected coverage measure* (ECM) of R with respect to μ.

For example, suppose information is recorded on $m + n$ files. A certain variable of interest is recorded on n files and missing on the remaining m. The variable is related by a regression model, (1.1) with $c_0 = 1$, to information recorded on all files. Over a period of time, some of the files are drawn at random. If the variable is missing, a prediction interval is required for its value. Suppose the loss arising from intervals failing to cover the true value is proportional to the number of such intervals. There is a second loss associated with the size of the intervals. It seems reasonable to construct the intervals according to

Received August 1985; revised February 1987.
[1]Research supported in part by the Natural Sciences and Engineering Research Council of Canada.
AMS 1980 *subject classifications.* Primary 62F25, 62J15; secondary 62A05, 62C20, 62H99, 62J10.
Key words and phrases. Confidence set, GMANOVA, Hunt–Stein theorem, invariant, minimax, multivariate regression, prediction region.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

criteria (1.2) and (1.3), with μ the uniform distribution on the standard basis for $\mathbb{R}^m$. The ECM is then the expected proportion of intervals that cover the true value. The choice of $1 - \alpha$ and w is discussed later.

Naiman (1984) considers estimation problems: In (1.1) take $c_0 = 0$, $C_2 = I_q$ and $\mathbf{y} = \beta$. One application is the construction of a confidence band for a polynomial regression curve. Let C_1 have ith row $(1, v_i, v_i^2, \ldots, v_i^{q-1})$. If μ is the distribution of $(1, \mathbf{v}, \mathbf{v}^2, \ldots, \mathbf{v}^{q-1})$, where $\mathbf{v}$ is distributed uniformly over a specified interval where the band is of interest, then the ECM is the expected proportion of the interval in which the true regression curve stays within the band. The idea generalizes easily to bounds for regression surfaces. Another application arises when estimating a finite number of linear functions, e.g., all simple contrasts in a one-way ANOVA. If μ is distributed uniformly on the set of linear functions of interest, then the ECM is the expected proportion of intervals covering the true value. Naiman minimizes (1.2) uniformly in (β, σ), subject to (1.3) and restrictions on the form of R. We strengthen his result in two ways: The restrictions on R are replaced by weaker invariance restrictions, proving that the optimal intervals are minimax, and the normality assumption is weakened, showing that the intervals are, in a certain sense, robust.

Consider weight functions of the form

$$w(a, \sigma) = w_1(a) w_2(a)/\sigma,$$

where $w_1(a) > 0$ and

$$w_2(a) = \left\{a'\left(c_0 I_m + C_2 (C_1'C_1)^{-1} C_2'\right)a\right\}^{-1/2}.$$

The $1/\sigma$ factor does not affect the form of the optimal invariant intervals but is included to make the minimax property nontrivial; cf. Hooper [(1982), page 1287]. The factor involving a is split into two terms for convenience. The theory of Section 2 shows that, for given μ, the following intervals are minimax:

$$|a'\mathbf{y} - a'C_2\hat{\beta}| \le \hat{\sigma} h(c, a)/w_2(a), \tag{1.4}$$

where

$$\hat{\beta} = (C_1'C_1)^{-1}C_1'\mathbf{x}, \qquad \hat{\sigma}^2 = (n - q)^{-1}\mathbf{x}'\left(I_n - C_1(C_1'C_1)^{-1}C_1'\right)\mathbf{x},$$

$$h(c, a) = \left\{(n - q)\left[\{c w_1(a)\}^{-2/(n-q+1)} - 1\right]^+\right\}^{1/2}.$$

Here $z^+ = z$ if $z > 0$ and $z^+ = 0$ if $z \le 0$. The constant c is chosen so that ECM $= 1 - \alpha$. Interchanging the order of intergration in (1.3) shows that

$$\text{ECM} = \int P\{a'\mathbf{y} \in R(a, \mathbf{x})\}\mu(da). \tag{1.5}$$

Since $\hat{\sigma}^{-1} w_2(a)(a'\mathbf{y} - a'C_2\hat{\beta})$ has a t distribution with $n - q$ degrees of freedom, it follows that c is the unique solution of

$$\int F(h(c, a))\mu(da) = 1 - \alpha/2,$$

where F is the t_{n-q} cumulative distribution function. Note that $F(h(\cdot, a))$ is continuous and decreasing over $\mathbb{R}^+$ and strictly decreasing when $h(\cdot, a) > 0$. An iterative method can be used to approximate c. Note that c depends on $1 - \alpha$, w_1 and μ.

The bound $1 - \alpha$ pertains to the relative importance of coverage and size in an overall average sense. The weight function, on the other hand, concerns the relative importance of coverage and size in a comparative sense, comparing among the various intervals. If all intervals are on the same footing, then $w_1 = 1/w_2$ seems natural. The choice $w_1 \equiv$ constant may be reasonable if the importance of coverage relative to size increases as $a'\mathbf{y}$ becomes harder to predict. If w_1 is constant, then $h(c, \cdot)$ is constant so the marginal coverage probability does not depend on a and, by (1.5), the ECM does not depend on μ. The Scheffé intervals [Miller (1981), page 48], the Tukey–Kramer intervals [Hayter (1984)] and the equal coverage Bonferroni intervals [Miller (1981), page 67] all have the form (1.4) with w_1 constant.

The ECM may be compared with the usual measure of coverage, the simultaneous coverage probability (SCP)

$$P\{a'\mathbf{y} \in R(a, \mathbf{x}) \text{ for all } a \in A\},$$

where A contains all linear functions of interest. If $\mu(A) = 1$, then

$$\text{SCP} \le \text{ECM}.$$

If A is a finite set containing N elements and μ is the uniform distribution on A, then the Bonferroni inequality gives

$$1 - N(1 - \text{ECM}) \le \text{SCP}.$$

The two coverage criteria reflect different loss functions. For further discussion, see Miller [(1981), pages 5–10 and 31–35] and Spjøtvoll (1972).

Section 2 presents the invariance theory in a general framework. Sections 3 and 4 contain applications to two multivariate generalizations of the preceding problem. Section 5 describes how optimal procedures can be modified to improve conditional performance.

There are applications where the linear functions of interest depend on the data $\mathbf{x}$. The criteria (1.2) and (1.3) are applicable if μ is changed to a conditional distribution $\mu(\cdot \mid \mathbf{x})$. This generalization is not treated here for two reasons: The distribution theory is less tractable and it is not clear what kind of conditional distributions are of practical interest. Conditional distributions are considered in the theoretical development of Section 2, but not in the applications.

2. Invariance theory. Let $\mathbf{x}$ and $\mathbf{y}$ be jointly distributed random variables taking values in $\mathscr{X}$ and $\mathscr{Y}$, with $\mathbf{y}$ currently unobservable. Let $\{\psi(\cdot; i, x): i \in \mathscr{I}, x \in \mathscr{X}\}$ be a family of functions defined on $\mathscr{Y}$ and taking values in a set $\mathscr{Z}$. Having observed $\mathbf{x} = x$, we wish to construct prediction regions for $\psi(\mathbf{y}; i, x)$, for various $i \in \mathscr{I}$. Let $\mathscr{P}_0$ be a family of probability measures on $\mathscr{I} \times \mathscr{X} \times \mathscr{Y}$ and suppose $(\mathbf{i}, \mathbf{x}, \mathbf{y})$ has distribution $P_0 \in \mathscr{P}_0$. The random variable $\mathbf{i}$ is introduced solely for notational convenience. In the preceding example, we have $\mathscr{I} = \mathbb{R}^m$, $i = a$, $\mathscr{Z} = \mathbb{R}$, $\psi(y; i, x) = a'y$, $\mathbf{i}$, $\mathbf{x}$ and $\mathbf{y}$ are independent and $\mathscr{L}(\mathbf{i}) = \mu$.

A *prediction function, or estimation function* if $\mathbf{y}$ is a parameter, is defined to be a measurable function φ: $\mathscr{I} \times \mathscr{X} \times \mathscr{Z} \to [0,1]$. The function φ determines prediction regions for $\psi(\mathbf{y}; i, x)$, $i \in \mathscr{I}$, via a randomizing mechanism as follows. Let $\mathbf{u}$ be distributed uniformly over $[0,1]$ and independently of $(\mathbf{i}, \mathbf{x}, \mathbf{y})$. Define

$$R_\varphi(i, x, u) = \{z \in \mathscr{Z}: u < \varphi(i, x, z)\}.$$

If the range of φ is contained in $\{0,1\}$, then φ is a *nonrandomized* prediction function. For given $P_0 \in \mathscr{P}_0$, the *expected coverage measure* of φ is defined to be

$$\begin{aligned} \mathrm{ECM}(\varphi: P_0) &= P\{\psi(\mathbf{y}; \mathbf{i}, \mathbf{x}) \in R_\varphi(\mathbf{i}, \mathbf{x}, \mathbf{u})\} \\ &= E\varphi(\mathbf{i}, \mathbf{x}, \psi(\mathbf{y}; \mathbf{i}, \mathbf{x})). \end{aligned} \tag{2.1}$$

Put $\mathbf{z} = \psi(\mathbf{y}; \mathbf{i}, \mathbf{x})$, $P = \mathscr{L}(\mathbf{i}, \mathbf{x}, \mathbf{z})$ and $\mathscr{P} = \{P: P_0 \in \mathscr{P}_0\}$. We then have

$$\mathrm{ECM}(\varphi; P_0) = E\varphi(\mathbf{i}, \mathbf{x}, \mathbf{z}) = \int \varphi \, dP.$$

The size of the region R_φ is specified by defining, for each $(i, x, P) \in \mathscr{I} \times \mathscr{X} \times \mathscr{P}$, a σ-finite measure $Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(dz|i, x)$ on $\mathscr{Z}$. In the preceding example, $Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(dz|i, x) = w(a, \sigma)\, dz$. Let Q_P be the σ-finite measure on $\mathscr{I} \times \mathscr{X} \times \mathscr{Z}$ given by

$$Q_P(di, dx, dz) = P^{\mathbf{i},\mathbf{x}}(di, dx) Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(dz|i, x), \tag{2.2}$$

where $P^{\mathbf{i},\mathbf{x}} = \mathscr{L}(\mathbf{i}, \mathbf{x})$. Note that

$$\int \varphi \, dQ_P = E\{Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(R_\varphi(\mathbf{i}, \mathbf{x}, \mathbf{u})|\mathbf{i}, \mathbf{x})\}.$$

Put $\mathscr{Q} = \{Q_P: P \in \mathscr{P}\}$. We say that φ_0 is *minimax at* ECM *level* $1 - \alpha$ for $(\mathscr{P}, \mathscr{Q})$ if φ_0 minimizes $\sup\{\int \varphi \, dQ: Q \in \mathscr{Q}\}$ subject to $\inf\{\int \varphi \, dP: P \in \mathscr{P}\} \geq 1 - \alpha$.

Put $\mathscr{V} = \mathscr{I} \times \mathscr{X} \times \mathscr{Z}$ and let $\mathscr{A}$ denote the product σ-field. Let G be a group acting on the left of $\mathscr{V}$. When defining group actions, it is convenient to start by considering transformations on $\mathscr{I} \times \mathscr{X} \times \mathscr{Y}$. Define, as usual, $gP(A) = P(g^{-1}A)$ and $gQ(A) = Q(g^{-1}A)$. In our applications, we have $gQ_P = Q_{gP}$. A prediction function φ is *invariant* if $\varphi(gv) = \varphi(v)$ for all $g \in G$ and $v \in \mathscr{V}$. We say that φ_0 is *optimal invariant at* ECM *level* $1 - \alpha$ for $(\mathscr{P}, \mathscr{Q})$ if, for each $Q \in \mathscr{Q}$, φ_0 minimizes $\int \varphi \, dQ$ among all invariant φ satisfying $\inf\{\int \varphi \, dP: P \in \mathscr{P}\} \geq 1 - \alpha$.

We obtain optimal invariant prediction functions by applying the following proposition. The proof is essentially an application of the Neyman–Pearson lemma and is omitted. For similar results, see Hooper [(1982), Theorem 1] and Takada [(1982), Theorem 2]. First note that if φ is invariant, then $\int \varphi \, d(gP) = \int \varphi \, dP$ and $\int \varphi \, d(gQ_P) = \int \varphi \, dQ_P$. Put $GP = \{gP: g \in G\}$ and $GQ_P = \{gQ_P: g \in G\}$.

Proposition 2.1. *Let* $\mathbf{t}$: $\mathscr{V} \to \mathscr{T}$ *be a maximal invariant under* G. *For fixed* P, *suppose* $P\mathbf{t}^{-1}$ *and* $Q_P\mathbf{t}^{-1}$ *admit densities* $p^{\mathbf{t}}$ *and* $q^{\mathbf{t}}$ *with respect to some* σ-*finite measure* λ *on* $\mathscr{T}$. *If* $\varphi = h(\mathbf{t})$ *is an invariant prediction function satisfy-*

ing $\int h(t)p^{t}(t)\lambda(dt) = 1 - \alpha$ *and*

$$h(t) = \begin{cases} 1, & \text{if } p^{t}(t) > cq^{t}(t), \\ 0, & \text{if } p^{t}(t) < cq^{t}(t), \end{cases}$$

then φ *is optimal invariant at ECM level* $1 - \alpha$ *for* (GP, GQ_P).

We need several regularity assumptions for the following version of the Hunt–Stein theorem. Suppose $\mathscr{A}$ is countably generated. Let $\mathscr{B}$ be a σ-field of subsets of G such that $Bg \in \mathscr{B}$ for each $B \in \mathscr{B}$, $g \in G$ and $\{(v, g): gv \in A\}$ is measurable $\mathscr{A} \times \mathscr{B}$ for each $A \in \mathscr{A}$. Suppose there exists a σ-finite measure ν over G such that $\nu(B) = 0$ implies $\nu(Bg) = 0$ for all $B \in \mathscr{B}$ and $g \in G$.

PROPOSITION 2.2. *Suppose G satisfies the Hunt–Stein condition: There exists an asymptotically right invariant sequence of probability distributions over $(G, \mathscr{B})$. Let μ_1 and μ_2 be σ-finite measures on $(\mathscr{V}, \mathscr{A})$. Given any measurable function φ: $\mathscr{V} \to [0,1]$, there exists an invariant measurable function φ_I: $\mathscr{V} \to [0,1]$ satisfying*

$$\inf_{g\in G} \int \varphi \, d(g\mu_i) \le \int \varphi_I \, d\mu_i \le \sup_{g\in G} \int \varphi \, d(g\mu_i)$$

for $i = 1, 2$.

Note that $G\mu_1$ and $G\mu_2$ need not be dominated families. Proposition 2.2 follows from a small modification in the proof of the Hunt–Stein theorem in Lehmann [(1986), page 519] together with an application of Theorem 4 in Lehmann [(1986), page 297].

3. Prediction in a multivariate regression model. We adopt the following notation: $\mathscr{M}(n, p)$ is the set of $n \times p$ matrices; $Gl(p)$ is the set of $p \times p$ invertible matrices; $G_U^+(p)$ is the set of $p \times p$ upper triangular matrices with positive diagonal elements; $\mathscr{O}(p)$ is the set of $p \times p$ orthogonal matrices; $\mathscr{S}(p)$ is the set of $p \times p$ positive definite symmetric matrices; $S^{1/2}$ is the symmetric square root of $S \in \mathscr{S}(p)$ and, for $n \ge p$, $\mathscr{F}_L(n, p)$ is the family of left $\mathscr{O}(n)$-invariant distributions on the set of $n \times p$ matrices of full rank; i.e., $\mathscr{L}(\mathbf{X}) \in \mathscr{F}_L(n, p)$ if $\mathscr{L}(\Gamma\mathbf{X}) = \mathscr{L}(\mathbf{X})$ for all $\Gamma \in \mathscr{O}(n)$ and $P\{\operatorname{rank}\mathbf{X} = p\} = 1$. Note that $\mathscr{F}_L(n, p)$ contains the multivariate normal distributions with mean 0 and covariance structure $I_n \otimes \Sigma$ for $\Sigma \in \mathscr{S}(p)$. Chmielewski (1981) reviews the literature on $\mathscr{F}_L(n, p)$.

Let $\mathbf{X}$ be a random $n \times p$ matrix. Partition $\mathbf{X}$ into $\mathbf{X}_{ij} \in \mathscr{M}(n_i, p_j)$, where $n_1 + n_2 = n$ and $p_1 + p_2 = p$. We put $\mathbf{x} = (\mathbf{X}_{11}, \mathbf{X}_{12}, \mathbf{X}_{22})$, $\mathbf{y} = \mathbf{X}_{21}$ and $\mathbf{z}' = \mathbf{a}'\mathbf{X}_{21}$, so $\mathbf{i} = \mathbf{a}$, $\mathscr{I} = \mathbb{R}^{n_2}$ and $\mathscr{Z} = \mathbb{R}^{p_1}$. Let $C \in \mathscr{M}(n, q)$ be a known matrix of constants with $C' = [C_1' : C_2']$, $C_i \in \mathscr{M}(n_i, q)$ and $\operatorname{rank} C_1 = q$. We assume that $n_1 \ge p + q$. Let $\mathscr{P}_0$ be the family of all distributions of $(\mathbf{i}, \mathbf{x}, \mathbf{y})$, or equivalently of $(\mathbf{a}, \mathbf{X})$, satisfying $\mathscr{L}(\mathbf{X} - CB) \in \mathscr{F}_L(n, p)$ for some $B \in \mathscr{M}(q, p)$, $\mathbf{X}$ and $\mathbf{a}$ are independent and $P\{\mathbf{a} = 0\} = 0$. The family $\mathscr{P}_0$ is indexed by $(\mathscr{L}(\mathbf{a}), B,$

$\mathscr{L}(\mathbf{X} - C B))$. Recall that $\mathscr{P} = \{\mathscr{L}(\mathbf{i}, \mathbf{x}, \mathbf{z}) \colon P_0 \in \mathscr{P}_0\}$. Put

$$
\begin{aligned}
&\nu_1 = p_1, \qquad \nu_2 = n_1 - p - q + 1,\\
&\hat{\mathbf{B}} = [\hat{\mathbf{B}}_1 : \hat{\mathbf{B}}_2] = (C_1'C_1)^{-1}C_1'[\mathbf{X}_{11} : \mathbf{X}_{12}],\\
&\mathbf{S} = (\mathbf{S}_{ij}) = [\mathbf{X}_{11} : \mathbf{X}_{12}]'\big(I_{n_1} - C_1(C_1'C_1)^{-1}C_1'\big)[\mathbf{X}_{11} : \mathbf{X}_{12}],\\
&\hat{\mathbf{X}}_{21} = C_2\hat{\mathbf{B}}_1 + (\mathbf{X}_{22} - C_2\hat{\mathbf{B}}_2)\mathbf{S}_{22}^{-1}\mathbf{S}_{21},\\
(3.1)\qquad &\mathbf{S}_{22} = \mathbf{U}_{22}\mathbf{U}_{22}', \qquad \mathbf{U}_{22} \in G_U^+(p_2),\\
&\mathbf{S}_{11\cdot 2} = \mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} = \mathbf{U}_{11\cdot 2}\mathbf{U}_{11\cdot 2}', \qquad \mathbf{U}_{11\cdot 2} \in G_U^+(p_1),\\
&\mathbf{W} = (\mathbf{X}_{22} - C_2\hat{\mathbf{B}}_2)\mathbf{U}_{22}'^{-1},\\
&\mathbf{v}' = (\mathbf{v}_1, \dots, \mathbf{v}_{p_1})\\
&\quad = \big\{\mathbf{a}'\big(I_{n_2} + C_2(C_1'C_1)^{-1}C_2' + \mathbf{W}\mathbf{W}'\big)\mathbf{a}\big\}^{-1/2}(\mathbf{z}' - \mathbf{a}'\hat{\mathbf{X}}_{21})\mathbf{U}_{11\cdot 2}'^{-1}.
\end{aligned}
$$

Lemma 3.2 shows that $\mathbf{W}$ is an ancillary statistic. We allow the weight function to depend on both $\mathbf{a}$ and $\mathbf{W}$. For $P \in \mathscr{P}$ with $E(\det \mathbf{U}_{11\cdot 2}) < \infty$, put

$$(3.2)\qquad Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(dz|a, W) = w_1(a, W)w_2(a, W)\{E(\det \mathbf{U}_{11\cdot 2})\}^{-1}\,dz,$$

where $w_1 > 0$ and

$$(3.3)\qquad w_2(a, W) = \big\{a'\big(I_{n_2} + C_2(C_1'C_1)^{-1}C_2' + WW'\big)a\big\}^{-p_1/2}.$$

Define Q_P by (2.2).

Now fix $\mathscr{L}(\mathbf{a}) \equiv \mu$. We will show that the optimal fully invariant ECM level $1 - \alpha$ prediction function φ_2 determines the following ellipsoidal regions for $z' = a'\mathbf{X}_{21}$:

$$(3.4)\quad (z' - a'\hat{\mathbf{X}}_{21})\mathbf{S}_{11\cdot 2}^{-1}(z' - a'\hat{\mathbf{X}}_{21})' \le \{w_2(a, \mathbf{W})\}^{-2/p_1}(\nu_1/\nu_2)h(c_2, a, \mathbf{W}),$$

where

$$h(c, a, W) = (\nu_2/\nu_1)\big[\{cw_1(a, W)\}^{-2/(\nu_1+\nu_2)} - 1\big]^+.$$

The constant c_2 is the solution of

$$(3.5)\qquad E\{F(h(c_2, \mathbf{a}, \mathbf{W}))\} = 1 - \alpha,$$

where F is the F_{ν_1,ν_2} cumulative distribution function. We also will show that a minimax ECM level $1 - \alpha$ prediction function φ_1 determines regions for $z' = a'\mathbf{X}_{21}$ via the inequality

$$(3.6)\qquad p^{\mathbf{v}}\big(\{w_2(a, \mathbf{W})\}^{1/p_1}(z' - a'\hat{\mathbf{X}}_{21})\mathbf{U}_{11\cdot 2}'^{-1}\big) \ge c_1 w_1(a, \mathbf{W}),$$

where $p^{\mathbf{v}}$ is the density of $\mathbf{v}$, given in Lemma 3.2, and c_1 is the α-quantile of

$$(3.7)\qquad \mathscr{L}(p^{\mathbf{v}}(\mathbf{v})/w_1(\mathbf{a}, \mathbf{W})).$$

Lemma 3.2 shows that c_1 and c_2 depend on $P \in \mathscr{P}$ only through $\mathscr{L}(\mathbf{a})$. Furthermore, if $w_1(a, W) = w_1(W)$, then c_1 and c_2 do not depend on $\mathscr{L}(\mathbf{a})$. A uniqueness argument shows that $\int \varphi_1\, dQ_P < \int \varphi_2\, dQ_P$ unless $p_1 = 1$, in which case $\varphi_1 = \varphi_2$; cf. Hooper [(1982), Correction]. However, numerical work in a related problem [Hooper and Yau, (1986), Table 1 with $n_1 = 1$] suggests that the improvement of φ_1 over φ_2 is likely to be slight unless ν_2 is small and p_1 large. One may judge that the greater complexity and the noninvariance of φ_1 offset this advantage.

Let G_2 be the group with elements $g = (\Gamma, A, F)$, where

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & I_{n_2} \end{pmatrix}, \qquad \Gamma_1 \in \mathscr{O}(n_1), \qquad \Gamma_1 C_1 = C_1,$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \qquad A_{11} \in Gl(p_1), \qquad A_{22} \in G_U^+(p_2)$$

and $F \in \mathscr{M}(q, p)$. The group operation is defined by the following actions on $\mathscr{I} \times \mathscr{X} \times \mathscr{Y}$ or, equivalently, $\mathbb{R}^{n_2} \times \mathscr{M}(n, p)$:

$$g(a, X) = (a, \Gamma X A' + CF).$$

The induced actions on $\mathscr{P}_0$ and $\mathscr{I} \times \mathscr{X} \times \mathscr{Z}$ are

$$g(\mathscr{L}(\mathbf{a}), B, \mathscr{L}(\mathbf{X} - CB)) = (\mathscr{L}(\mathbf{a}), BA' + F, \mathscr{L}((\mathbf{X} - CB)A')),$$

$$\begin{aligned} g(a, [X_{11} : X_{12}], X_{22}, z') = (a, \Gamma_1[X_{11} : X_{12}]A' + C_1 F, X_{22}A'_{22} \\ + C_2 F_2, z'A'_{11} + a'(X_{22}A'_{12} + C_2 F_1)), \end{aligned}$$

where $F = [F_1 : F_2]$, $F_i \in \mathscr{M}(q, p_i)$. Let G_1 denote the subgroup of G_2 with $A \in G_U^+(p)$.

The translations $X \to X + CF$ and the linear transformations of variables $X \to XA'$ are the important group actions. The orthogonal transformations $X \to \Gamma X$ are introduced for convenience; they do not affect the form of the optimal regions. The orthogonal transformations preserve $\mathscr{L}(\mathbf{X})$ so the corresponding reductions could be obtained by sufficiency. To justify the sufficiency-invariance route one would need to verify a regularity condition similar to Assumption B of Hall, Wijsman and Ghosh [(1965), page 605].

LEMMA 3.1. *Maximal invariants on $\mathscr{I} \times \mathscr{X} \times \mathscr{Z}$ under G_1 and G_2 are given by* $(\mathbf{a}, \mathbf{W}, \mathbf{v})$ *and* $(\mathbf{a}, \mathbf{W}, \|\mathbf{v}\|^2)$, *respectively.*

The proof uses standard arguments and is omitted.

LEMMA 3.2. *The variables* $\mathbf{a}$, $\mathbf{W}$ *and* $\mathbf{v}$ *are independent and* $\mathscr{L}(\mathbf{W}, \mathbf{v})$ *is the same for all* $P \in \mathscr{P}$. *Furthermore,* $(\nu_2/\nu_1)\|\mathbf{v}\|^2 \sim F_{\nu_1, \nu_2}$ *and* $\mathbf{v}$ *has the following density with respect to Lebesgue measure*:

$$\begin{aligned} p^{\mathbf{v}}(v) = \pi^{-p_1/2} \prod_{i=1}^{p_1} [\Gamma((n_1 - q - p_2 - i + 2)/2)/\Gamma((n_1 - q - p_2 - i + 1)/2)] \\ \times (1 + \|v\|^2)^{-(n_1 - p - q)/2} \prod_{i=1}^{p_1} \left(1 + v_i^2 + v_{i+1}^2 + \cdots + v_{p_1}^2\right)^{-1}. \end{aligned}$$

PROOF. By assumption, $\mathbf{a}$ and $\mathbf{X}$ are independent. Fixing $a \in R^{n_2} - \{0\}$, we derive the conditional distribution of $(\mathbf{v}, \mathbf{W})$ given $\mathbf{a} = a$. By Corollary 1.1 of Kariya (1981), the distribution of any G_1-invariant function of $\mathbf{X}$ is the same for all $P_0 \in \mathscr{P}_0$. Thus we may assume, without loss of generality, that $\mathbf{X} \sim N(0, I_n \otimes I_p)$.

We have $\hat{\mathbf{B}}$, $\mathbf{S}$, $\mathbf{X}_{21}$ and $\mathbf{X}_{22}$ independent with $\hat{\mathbf{B}} \sim N_{q\times p}(0, (C_1'C_1)^{-1} \otimes I_p)$ and $\mathbf{S} \sim W_p(n_1 - q, I_p)$. Putting

$$\mathbf{H} = [\mathbf{H}_1 : \mathbf{H}_2] = [\mathbf{X}_{21} : \mathbf{X}_{22}] - C_2\hat{\mathbf{B}},$$

we have $\mathbf{H}$ and $\mathbf{S}$ independent and

$$\mathbf{H} \sim N_{n_2\times p}\Big(0, \big(I_{n_2} + C_2(C_1'C_1)^{-1}C_2'\big) \otimes I_p\Big).$$

Observe that $\mathbf{W} = \mathbf{H}_2\mathbf{U}_{22}'^{-1}$ so $\mathbf{W}\mathbf{W}' = \mathbf{H}_2\mathbf{S}_{22}^{-1}\mathbf{H}_2'$. A standard argument shows that, given $(\mathbf{H}_2, \mathbf{S}_{22}) = (H_2, S_{22})$, the conditional distribution of $\mathbf{H}_{1\cdot 2} \equiv \mathbf{H}_1 - \mathbf{H}_2\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ is

$$\mathscr{L}(\mathbf{H}_{1\cdot 2}|H_2, S_{22}) = N_{n_2\times p_1}\Big(0, \big(I_{n_2} + C_2(C_1'C_1)^{-1}C_2' + WW'\big) \otimes I_{p_1}\Big)$$

and also that $\mathbf{S}_{11\cdot 2}$ is independent of $(\mathbf{H}_{1\cdot 2}, \mathbf{W})$. Put

$$\mathbf{h}' = \Big\{\mathbf{a}'\big(I_{n_2} + C_2(C_1'C_1)^{-1}C_2' + \mathbf{W}\mathbf{W}'\big)\mathbf{a}\Big\}^{-1/2}\mathbf{a}'\mathbf{H}_{1\cdot 2}.$$

It follows that $\mathbf{h}$, $\mathbf{S}_{11\cdot 2}$ and $\mathbf{W}$ are independent with $\mathbf{h} \sim N_{p_1}(0, I_{p_1})$ and $\mathbf{S}_{11\cdot 2} \sim W_{p_1}(n_1 - q - p_2, I_{p_1})$. Noting that $\mathbf{v}' = \mathbf{h}'\mathbf{U}_{11\cdot 2}'^{-1}$, we see that $(n_1 - q - p_2)\|\mathbf{v}\|^2$ has Hotelling's T^2 distribution. The density of $\mathbf{v}$ is obtained from Theorem 4.2(ii) of Olkin and Rubin (1964). Their constant term c contains an error: Replace 2π by π.

We have shown that $\mathbf{v}$ and $\mathbf{W}$ are conditionally independent given $\mathbf{a}$ and that $\mathbf{v}$ and $\mathbf{a}$ are independent. Since $\mathbf{W}$ and $\mathbf{a}$ are independent, it follows that $\mathbf{v}$, $\mathbf{W}$ and $\mathbf{a}$ are independent. □

LEMMA 3.3. *For all $P \in \mathscr{P}$ with $E(\det \mathbf{U}_{11\cdot 2}) < \infty$, we have*

$$E(\det \mathbf{U}_{11\cdot 2}|\mathbf{W}) = E(\det \mathbf{U}_{11\cdot 2}) \qquad a.s.$$

PROOF. Since the joint distribution $(\mathbf{U}_{11\cdot 2}, \mathbf{W})$ does not involve B, we assume, without loss of generality, that $B = 0$; i.e., $\mathscr{L}(\mathbf{X}) \in \mathscr{F}_L(n, p)$. By Proposition 7.3 of Eaton (1983), we have $\mathbf{X} = \mathbf{Z}\mathbf{A}'$, where $\mathbf{Z}$ and $\mathbf{A}$ are independent, $\mathbf{Z}$ has the uniform distribution on the Stiefel manifold and $\mathbf{A} \in G_U^+(p)$. Writing $\mathbf{U}_{11\cdot 2} = \mathbf{U}_{11\cdot 2}(\mathbf{X})$ and $\mathbf{W} = \mathbf{W}(\mathbf{X})$, we observe that $\mathbf{U}_{11\cdot 2}(XA') = A_{11}\mathbf{U}_{11\cdot 2}(X)$ and $\mathbf{W}(XA') = \mathbf{W}(X)$ for all $X \in \mathscr{M}(n, p)$ and $A \in G_U^+(p)$. Thus, we have

$$\begin{aligned} E(\det \mathbf{U}_{11\cdot 2}|\mathbf{W}) &= E\{\det \mathbf{U}_{11\cdot 2}(\mathbf{Z}\mathbf{A}')|\mathbf{W}(\mathbf{Z}\mathbf{A}')\} \\ &= E\{(\det \mathbf{A}_{11})(\det \mathbf{U}_{11\cdot 2}(\mathbf{Z}))|\mathbf{W}(\mathbf{Z})\} \\ &= E\{\det \mathbf{A}_{11}\}E\{\det \mathbf{U}_{11\cdot 2}(\mathbf{Z})|\mathbf{W}(\mathbf{Z})\} \quad \text{a.s.} \end{aligned} \tag{3.8}$$

A similar argument yields

$$E(\det \mathbf{U}_{11\cdot 2}) = E(\det \mathbf{A}_{11})E\{\det \mathbf{U}_{11\cdot 2}(\mathbf{Z})\}. \tag{3.9}$$

Now suppose $\mathscr{L}(\mathbf{X}) = N(0, I_n \otimes I_p)$. In this case, $\mathbf{U}_{11\cdot 2}$ and $\mathbf{W}$ are independent, so (3.8) and (3.9) are equal a.s. This shows that

$$E\{\det \mathbf{U}_{11\cdot 2}(\mathbf{Z})|\mathbf{W}(\mathbf{Z})\} = E\{\det \mathbf{U}_{11\cdot 2}(\mathbf{Z})\} \quad \text{a.s.} \tag{3.10}$$

Combining (3.8)–(3.10) yields the desired result for general $\mathscr{L}(\mathbf{X}) \in \mathscr{F}_L(n, p)$. $\square$

THEOREM 3.1. *Fix $P \in \mathscr{P}$ with $E(\det \mathbf{U}_{11\cdot 2}) < \infty$, and let φ_1 and φ_2 be the indicator functions determined by* (3.6) *and* (3.4).

(i) *φ_1 is optimal G_1-invariant and minimax at ECM level $1 - \alpha$ for (G_1P, G_1Q_P);*
(ii) *φ_2 is optimal G_2-invariant at ECM level $1 - \alpha$ for (G_2P, G_2Q_P).*

PROOF. We first show φ_1 is optimal G_1-invariant by applying Proposition 2.1 with $\mathbf{t} = (\mathbf{a}, \mathbf{W}, \mathbf{v})$ and $\lambda(dt) = P^{\mathbf{a}}(da)P^{\mathbf{W}}(dW)\,dv$. By Lemma 3.2 we have $p^{\mathbf{t}}(t) = p^{\mathbf{v}}(v)$. To determine $q^{\mathbf{t}}$, we condition on $(\mathbf{a}, \mathbf{x})$, compute the Jacobian $|\det dz/dv|$ and then take the conditional expectation given $(\mathbf{a}, \mathbf{W})$. In the last step, Lemma 3.3 and the independence of $\mathbf{a}$ and $(\mathbf{U}_{11\cdot 2}, \mathbf{W})$ show that $E(\det \mathbf{U}_{11\cdot 2}|\mathbf{a}, \mathbf{W}) = E(\det \mathbf{U}_{11\cdot 2})$ a.s. We obtain $q^{\mathbf{t}}(t) = w_1(a, W)$. Minimaxity of φ_1 follows from Proposition 2.2 and the fact that G_1 satisfies the Hunt–Stein condition; see Bondar and Milnes [(1981), Section 2].

Part (ii) follows from Proposition 2.1 with $\mathbf{t} = (\mathbf{a}, \mathbf{W}, \mathbf{t}_1)$, $\mathbf{t}_1 = \|\mathbf{v}\|^2$ and $\lambda(dt) = P^{\mathbf{a}}(da)P^{\mathbf{W}}(dW)\,dt_1$. We compute $p^{\mathbf{t}}(t) \propto t_1^{(\nu_1 - 2)/2}(1 + t_1)^{-(\nu_1 + \nu_2)/2}$ and $q^{\mathbf{t}}(t) \propto w_1(a, W)t_1^{(\nu_1 - 2)/2}$. $\square$

4. Estimation in GMANOVA. Kariya (1978, 1981, 1985), Marden (1983) and Kariya and Sinha (1985) present optimality results for the general multivariate analysis of variance (GMANOVA) hypothesis testing problem. Hooper (1983) and Hooper and Yau (1986) consider GMANOVA confidence estimation problems. We present the GMANOVA model in canonical form. Let $\mathbf{X}$ be a random $n \times p$ matrix partitioned into $\mathbf{X}_{ij} \in \mathscr{M}(n_i, p_j)$, where $n_1 + n_2 + n_3 = n$ and $p_1 + p_2 + p_3 = p$. We assume that $n \geq p$ and $n_3 \geq p_2 + p_3$. Suppose $\mathscr{L}(\mathbf{X} - M) \in \mathscr{F}_L(n, p)$, with M partitioned in the same way as $\mathbf{X}$. We consider simultaneous estimation of $a'M_{12}b$ for some vectors $a \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{p_2}$, so here we have $\mathbf{x} = \mathbf{X}$, $\mathscr{X} = \mathscr{M}(n, p)$, $\mathbf{y} = M_{12}$, $\mathscr{Y} = \mathscr{M}(n_1, p_2)$, $\mathbf{i} = (\mathbf{a}, \mathbf{b})$, $\mathscr{I} = \mathbb{R}^{n_1} \times \mathbb{R}^{p_2}$, $\mathbf{z} = \mathbf{a}'M_{12}\mathbf{b}$ and $\mathscr{Z} = \mathbb{R}$.

Put

$$
\begin{aligned}
\mathbf{S} &= (\mathbf{S}_{ij}) = [\mathbf{X}_{31} : \mathbf{X}_{32} : \mathbf{X}_{33}]'[\mathbf{X}_{31} : \mathbf{X}_{32} : \mathbf{X}_{33}],\\
\hat{\mathbf{M}}_{12} &= \mathbf{X}_{12} - \mathbf{X}_{13}\mathbf{S}_{33}^{-1}\mathbf{S}_{32},\\
\mathbf{S}_{33} &= \mathbf{U}_{33}\mathbf{U}_{33}', \qquad \mathbf{U}_{33} \in G_U^+(p_3),\\
\mathbf{S}_{22\cdot 3} &= \mathbf{S}_{22} - \mathbf{S}_{23}\mathbf{S}_{33}^{-1}\mathbf{S}_{32},\\
\mathbf{W} &= \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{13} \\ \mathbf{X}_{23} \end{bmatrix} \mathbf{U}_{33}'^{-1},\\
\mathbf{T}_2 &= \mathbf{W}_1\mathbf{W}_1' = \mathbf{X}_{13}\mathbf{S}_{33}^{-1}\mathbf{X}_{13}',\\
\mathbf{v} &= \{\mathbf{a}'(I_{n_1} + \mathbf{T}_2)\mathbf{a}\}^{-1/2}\{\mathbf{b}'\mathbf{S}_{22\cdot 3}\mathbf{b}/(n_3 - p_3)\}^{-1/2}(\mathbf{z} - \mathbf{a}'\hat{\mathbf{M}}_{12}\mathbf{b}).
\end{aligned}
\tag{4.1}
$$

Let $\mathscr{P}_0$ be the family of all distributions of $(\mathbf{i}, \mathbf{x}, \mathbf{y})$ satisfying $\mathscr{L}(\mathbf{X} - M) \in \mathscr{F}_L(n, p)$ for some $M \in \mathscr{M}(n, p)$ with $M_{ij} = 0$ if $i = 3$ or $j = 3$, $P\{\mathbf{y} = M_{12}\} = 1$, $(\mathbf{a}, \mathbf{b})$ and $\mathbf{X}$ are independent, $P\{\mathbf{a} = 0\} = P\{\mathbf{b} = 0\} = 0$ and $E[\{\mathbf{b}'\mathbf{S}_{22\cdot 3}\mathbf{b}\}^{1/2}|\mathbf{b}] < \infty$ a.s. The last restriction ensures the existence of confidence intervals with finite expected length. The family $\mathscr{P}_0$ is indexed by $(\mathscr{L}(\mathbf{a}, \mathbf{b}), M, \mathscr{L}(\mathbf{X} - M))$. Let $\mathscr{P} = \{\mathscr{L}(\mathbf{i}, \mathbf{x}, \mathbf{z})\colon P_0 \in \mathscr{P}_0\}$.

Lemma 4.1 shows that $\mathbf{W}$ is an ancillary statistic. We consider weight functions based on $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{W}$. Put

$$Q_P^{\mathbf{z}|\mathbf{i},\mathbf{x}}(dz|a, b, W) = w_1(a, b, W)\{a'(I_{n_1} + T_2)a\}^{-1/2} \times \left[E\{b'\mathbf{S}_{22\cdot 3}b\}^{1/2}\right]^{-1} dz \tag{4.2}$$

and define Q_P by (2.2). Let φ_0 be the estimation function that determines the intervals for $z = a'M_{12}b$,

$$|z - a'\hat{\mathbf{M}}_{12}b| \le \{a'(I_{n_1} + \mathbf{T}_2)a\}^{1/2}\{(b'\mathbf{S}_{22\cdot 3}b)/(n_3 - p_3)\}^{1/2} \times h(c_0, a, b, \mathbf{W}), \tag{4.3}$$

where

$$h(c, a, b, W) = \left\{(n_3 - p_3)\left[\{cw_1(a, b, W)\}^{-2/(n_3 - p_3 + 1)} - 1\right]^+\right\}^{1/2}.$$

The constant c_0 is the solution of

$$E\{F(h(c_0, \mathbf{a}, \mathbf{b}, \mathbf{W}))\} = 1 - \alpha/2, \tag{4.4}$$

where F is the $t_{n_3 - p_3}$ cumulative distribution function. We note that c_0, and hence φ_0, depends on $1 - \alpha$, w_1 and $\mathscr{L}(\mathbf{a}, \mathbf{b})$. If $w_1 = w_1(W)$, then $h = h(c, W)$, the marginal coverage probability of (4.3) is $1 - \alpha$ for all a, b and the ECM is $1 - \alpha$ for all μ. We will show that φ_0 possesses certain optimality properties.

Lemma 4.1. (i) *The random variables* $\mathbf{v}$, $\mathbf{W}$ *and* $(\mathbf{a}, \mathbf{b})$ *are independent and* $\mathscr{L}(\mathbf{v}, \mathbf{W})$ *is the same for all* $P \in \mathscr{P}$. *Furthermore* $\mathbf{v} \sim t_{n_3 - p_3}$.

(ii) *For all* $P \in \mathscr{P}$ *we have*

$$E\left[\{\mathbf{b}'\mathbf{S}_{22\cdot 3}\mathbf{b}\}^{1/2}|\mathbf{a}, \mathbf{b}, \mathbf{W}\right] = E\left[\{\mathbf{b}'\mathbf{S}_{22\cdot 3}\mathbf{b}\}^{1/2}|\mathbf{b}\right] \quad a.s.$$

Proof. For (i) we first claim that the conditional distribution of $(\mathbf{v}, \mathbf{W})$ given $(\mathbf{a}, \mathbf{b}) = (a, b)$ is the same for all $P \in \mathscr{P}$, $a \ne 0$, $b \ne 0$. Put $\mathbf{Y}_{i2} = (\mathbf{X}_{i2} - M_{i2})b$, $\mathbf{Y}_{i3} = \mathbf{X}_{i3}$ and $\mathbf{Y} = (Y_{ij}) \in \mathbf{M}(n, 1 + p_3)$. It is easy to check that $(\mathbf{v}, \mathbf{W})$ is a function of $\mathbf{Y}$ that is invariant under the actions $\mathbf{Y} \to \mathbf{Y}A'$, $A \in G_U^+(1 + p_3)$. Also note that $\mathscr{L}(\mathbf{Y}) \in \mathscr{F}_L(n, p)$. The claim follows from Corollary 1.1 of Kariya (1981). Thus we may assume, without loss of generality, that $\mathscr{L}(\mathbf{X} - M) = N(0, I_n \otimes I_p)$. The rest of the proof of (i) is straightforward; see Kariya [(1978), Lemma 3.1].

For (ii) we observe that $\{b'\mathbf{S}_{22\cdot 3}b\}^{1/2} = f(\mathbf{Y})$, where

$$f(Y) = \left\{Y_{32}'\left(I_{n_3} - Y_{33}(Y_{33}'Y_{33})^{-1}Y_{33}'\right)Y_{32}\right\}^{1/2}$$

satisfies $f(YA') = a_{11} f(Y)$ for all $A = (a_{ij}) \in G_U^+(1 + p_3)$. The rest of the proof of (ii) is similar to that of Lemma 3.3. □

For each distribution μ on $\mathscr{I} = \mathbb{R}^{n_1} \times \mathbb{R}^{p_2}$, put $\mathscr{P}_\mu = \{P \in \mathscr{P}: \mathscr{L}(\mathbf{a}, \mathbf{b}) = \mu\}$.

THEOREM 4.1. *Fix $1 - \alpha$, $\mu = \mathscr{L}(\mathbf{a}, \mathbf{b})$ and w_1. For each $P \in \mathscr{P}_\mu$, φ_0 minimizes $\int \varphi \, dQ_P$ among all estimation functions based on $(\mathbf{a}, \mathbf{b}, \mathbf{v}, \mathbf{W})$ satisfying $\int \varphi \, dP \geq 1 - \alpha$.*

PROOF. This is essentially an application of the Neyman–Pearson lemma, as in Proposition 2.1. We take $\mathbf{t} = (\mathbf{a}, \mathbf{b}, \mathbf{W}, \mathbf{v})$, $\lambda(dt) = \mu(da, db) P^{\mathbf{W}}(dW)\, dv$ and applying Lemma 4.1(i) obtain

$$p^{\mathbf{t}}(t) = p^{\mathbf{v}}(v) \propto \left\{1 + v^2/(n_3 - p_3)\right\}^{-(n_3 - p_3 + 1)/2}.$$

To determine $q^{\mathbf{t}}$ we condition on $(\mathbf{a}, \mathbf{b}, \mathbf{x})$, compute the Jacobian $|dz/dv|$ and then take the conditional expectation given $(\mathbf{a}, \mathbf{b}, \mathbf{W})$. Applying Lemma 4.1(ii) yields $q^{\mathbf{t}}(t) \propto w_1(a, b, W)$. □

The following minimaxity result is applicable when intervals are desired for $a'M_{12}b$ with b restricted to a given set of linearly independent vectors. I do not know whether φ_0 is minimax for general $\mathscr{L}(\mathbf{a}, \mathbf{b})$. Put $\mathscr{P}_N = \{P \in \mathscr{P}: \mathscr{L}(\mathbf{X})$ is multivariate normal$\}$.

THEOREM 4.2. *Fix $1 - \alpha$, $\mu = \mathscr{L}(\mathbf{a}, \mathbf{b})$ and w_1. Suppose $P\{\mathbf{b} \in B\} = 1$ for some basis B of $\mathbb{R}^{p_2}$. Let $\mathscr{P}_1$ be a family of distributions with $\mathscr{P}_\mu \cap \mathscr{P}_N \subseteq \mathscr{P}_1 \subseteq \mathscr{P}_\mu$. Put $\mathscr{Q}_1 = \{Q_P: P \in \mathscr{P}_1\}$. The estimation function φ_0 is minimax at ECM level $1 - \alpha$ for $(\mathscr{P}_1, \mathscr{Q}_1)$.*

PROOF. Consider a transformation $X \to XA'$, $M_{12} \to M_{12}A'_{22}$, $b \to A'^{-1}_{22} b$, where $A = (A_{ij})$ is block diagonal. We may choose A_{22} so that $\{A'^{-1}_{22} b: b \in B\}$ is the standard basis. Thus without loss of generality we assume that B is the standard basis.

Let P be the distribution in $\mathscr{P}_\mu$ with $\mathscr{L}(\mathbf{X}) = N(0, I_n \otimes I_p)$. Let G be the group with elements $g = (\Gamma, A, F)$, where

$$\Gamma = \begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & \Gamma_3 \end{bmatrix}, \qquad A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}, \qquad F = \begin{bmatrix} F_{11} & F_{12} & 0 \\ F_{21} & F_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$\Gamma_3 \in \mathscr{O}(n_3)$, $A \in G_U^+(p)$, A_{22} is diagonal and $F \in \mathscr{M}(n, p)$. The group operation is defined by the following actions on $\mathscr{I} \times \mathscr{X}$:

$$g(a, b, X) = (a, b, \Gamma X A' + F).$$

The induced action on $\mathscr{Y}$ is $gM_{12} = M_{12}A'_{22} + F_{12}$. Note that $M_{12}A'_{22}b = (b'A_{22}b)M_{12}b$ for A_{22} diagonal and $b \in B$. Thus there is an action induced on

$\mathbb{R}^{n_1} \times B \times \mathscr{Z}$,

$$g(a, b, z) = (a, b, (b'A_{22}B)z + a'F_{12}b).$$

Let $\mathbf{R}$ be the correlation matrix obtained from $\mathbf{S}_{22 \cdot 3}$. A standard argument shows that $\mathbf{t} = (\mathbf{a}, \mathbf{b}, \mathbf{W}, \mathbf{R}, \mathbf{v})$ is a maximal invariant on $\mathbb{R}^{n_1} \times B \times \mathscr{X} \times \mathscr{Z}$. An application of Proposition 2.1 with

$$\lambda(dt) = \mu(da, db) P^{\mathbf{W}}(dW) P^{\mathbf{R}}(dR)\, dv$$

shows that φ_0 is optimal invariant at ECM level $1 - \alpha$ for (GP, GQ_P). Here we use the fact that, under GP, $\mathbf{R}$ is ancillary and independent of $(b'\mathbf{S}_{22 \cdot 3}b\colon b \in B)$.

The group G satisfies the Hunt–Stein condition; see Bondar and Milnes [(1981), Section 2]. Hence φ_0 is minimax at ECM level $1 - \alpha$ for (GP, GQ_P) by Proposition 2.2. Finally we observe that

$$\int \varphi_0 \, dQ_P = E\{w_1(\mathbf{a}, \mathbf{b}, \mathbf{W}) h(c, \mathbf{a}, \mathbf{b}, \mathbf{W}) / (n_3 - p_3)^{1/2}\}$$

is the same for all $P \in \mathscr{P}_\mu$. Since $GP \subseteq \mathscr{P}_1 \subseteq \mathscr{P}_\mu$, the conclusion of the theorem follows. $\square$

5. Conditionality properties. Conditioning is of interest in two respects. First, the ECM is a weighted average of the marginal coverage probabilities: $\text{ECM} = E\varphi = E\{E(\varphi|\mathbf{i})\}$. It seems desirable to report the marginal coverage probabilities $E(\varphi|i)$ to avoid leaving the impression that these are equal when this is not the case. Second, one may wish to condition on the ancillary statistic $\mathbf{W}$; i.e., solve the optimization problem with $\mathbf{W} = W$ fixed. This leads to regions defined as before except that $\mathbf{W}$ is replaced by the nonrandom W in (3.5), (3.7) and (4.4). Thus the constants c_i are replaced by functions $c_i(W)$ defined so that $E(\varphi|W) = 1 - \alpha$, where φ is given by (3.4), (3.6) and (4.3). Note that the calculation of $c_i(W)$ is easier than the calculation of c_i in the unconditional problem. The conditional solutions have $E(\varphi|i) = E(\varphi|i, W)$.

The following is a brief description of some theory supporting the above modifications. Robinson (1979) studies conditional properties of statistical procedures on the basis of their ability to withstand betting procedures. This idea may be applied in the present context simply by replacing $(\mathscr{X}, \mathscr{L}(\mathbf{x}))$ with $(\mathscr{I} \times \mathscr{X}, \mathscr{L}(\mathbf{i}, \mathbf{x}))$ in the definition of a betting procedure. Under this modification, Hooper (1984) shows that the conditional versions of (3.4) and (3.6) do not admit superrelevant betting procedures provided the marginal coverage probabilities are reported. The same is true for the conditional version of (4.3) under the additional assumption that $\mathbf{b}$ is restricted to a basis. I do not know whether the result is true for general $\mathscr{L}(\mathbf{a}, \mathbf{b})$.

Acknowledgments. I wish to thank the Department of Statistics, Stanford University, for its hospitality during part of the preparation of this article. I am grateful to the Associate Editor and referees for their comments on the presentation of the material.

REFERENCES

BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. verw. Gebiete* **57** 103–128.

CHMIELEWSKI, M. A. (1981). Elliptically symmetric distributions: A review and bibliography. *Internat. Statist. Rev.* **49** 67–74.

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

HALL, W. J., WIJSMAN, R. A. and GHOSH, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

HAYTER, A. J. (1984). A proof of the conjecture that the Tukey–Kramer multiple comparison procedure is conservative. *Ann. Statist.* **12** 61–75.

HOOPER, P. M. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10** 1283–1294. Correction **12** (1984) 784.

HOOPER, P. M. (1983). Simultaneous interval estimation in the general multivariate analysis of variance model. *Ann. Statist.* **11** 666–673. Correction **12** (1984) 785.

HOOPER, P. M. (1984). A conditional property of invariant confidence and prediction regions. *Ann. Statist.* **12** 745–750.

HOOPER, P. M. and YAU, W. K. (1986). Optimal confidence regions in GMANOVA. *Canad. J. Statist.* **14** 315–322.

KARIYA, T. (1978). The general MANOVA problem. *Ann. Statist.* **6** 200–214.

KARIYA, T. (1981). Robustness of multivariate tests. *Ann. Statist.* **9** 1267–1275.

KARIYA, T. (1985). *Testing in the Multivariate General Linear Model*. Kinokuniya, Tokyo.

KARIYA, T. and SINHA, B. K. (1985). Nonnull and optimality robustness of some tests. *Ann. Statist.* **13** 1182–1197.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.

MARDEN, J. I. (1983). Admissibility of invariant tests in the general multivariate analysis of variance problem. *Ann. Statist.* **11** 1086–1099.

MILLER, R. G. (1981). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York.

NAIMAN, D. Q. (1984). Optimal simultaneous confidence bounds. *Ann. Statist.* **12** 702–715.

OLKIN, I. and RUBIN, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* **35** 261–269.

ROBINSON, G. K. (1979). Conditional properties of statistical procedures. *Ann. Statist.* **7** 742–755.

SPJØTVOLL, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.* **43** 398–411.

TAKADA, Y. (1982). Application of an adequate statistic to the invariant prediction region. *Ann. Inst. Statist. Math.* **34** 491–503.

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA T6G 2G1
CANADA